newly diagnosed malignant glioma. Gliadel treatment produced a risk reduction of 63 percent as shown here, and the intent to treat population with 95 percent confidence intervals of 17 to 82 percent. The trial was positive in the pre-specified efficacy endpoints in the overall intent to treat population as well as in the GBM subpopulation when one accounts for known important prognostic factors in addition to tumor histology; those being age and Karnofsky performance status.

Gliadel was well tolerated in this trial. However, only 16 patients with primary malignant glioma were treated with Gliadel wafers in this trial. Thus, a larger study was necessary to better define the safety in a clinical setting and to provide a more precise estimate of benefit and efficacy.

I'd like now to proceed to the T-301 study. Now the objectives of this second Phase III study were identical to the 0190 study. That is to say it studied the efficacy and safety of Gliadel wafers versus the placebo wafers when used in conjunction with surgery and radiotherapy to prolong survival in patients with newly diagnosed malignant glioma. The reasons for the T-301 study

included the desire to have a larger safety sample of Gliadel wafer treated patients in the primary surgery setting where the patients will receive radiotherapy shortly after the implantation of Gliadel wafer, and to confirm the clinical benefit of Gliadel wafer treatment.

The key points of the design of the T-301 study are shown here on this slide. The trial was a randomized, double-blind, placebo-controlled study. The primary pre-specified efficacy endpoint was overall survival in all patients randomized, the ITT populations, pre-specified primary endpoint by the Kaplan-Meier method 12 months after the final patient was enrolled. Therefore, some patients had a longer period of follow-up than others, but every patient had 12 months of follow-up. The T-301 study design protocol and statistical analysis plan were provided to the FDA in advance of completing patient follow-up and any unblinding of data.

Pre-specified secondary efficacy endpoints in this trial shown here included overall survival in the GBM subpopulation of patients as well as a number of important clinical endpoints which include time to Karnofsky performance decline, time

to neuroperformance decline, progression-free survival, and a quality of life evaluation.

Now the T-301 study was predominantly a European study. There were 42 sites, as shown here, in 14 different countries, including the United States, that recruited patients. All centers that enrolled patients in this trial were regional centers of excellence with active brain tumor surgery services. The major inclusion criteria for patients is shown here on this slide. They're identical to the 0190 study that I briefly reviewed and similar to other trials in this patient population of primary malignant glioma.

Male and female patients ages 18 to 65 were enrolled. Patients could only have a single contrast enhancing unilateral lesion diagnosed by cranial MRI or CT scan. Surgical treatment was provided within two weeks of the baseline scan. And patients had to have a Karnofsky performance score of 60 or higher. And they could not have had previous treatment for the suspected diagnosis of primary malignant glioma.

Now 240 patients were enrolled in this study with 120 patients in each treatment group as shown here. The baseline characteristics of the

group are shown on this -- some of the baseline characteristics are shown on this slide. The mean age and range of the two treatment groups were similar as was sex distribution. In addition, the tumor types were very similarly distributed between these two treatment groups with the GBM subtype here representing about 80 to 85 percent of both treatment groups. As I previously noted, age and tumor histology are known prognostic factors that influence survival.

Now another important baseline characteristic that's an impact on patient survival is Karnofsky score. There were no significant differences between the treatment groups in the baseline Karnofsky score, however there were more patients in the Gliadel group with lower Karnofsky performance scores, which would be expected to confer a worse prognosis.

An additional baseline characteristic that may influence survival is tumor volume. The Gliadel wafer treatment group had a significantly larger tumor, shown here, than the placebo wafer treated group, p less than 0.05, although there were a significant number of missing data in this value. The percentage of tumor resected did not

differ between the two treatment groups.

Now before presenting the efficacy and safety results of the T-301 study I think it's important to address a number of statistical analytic and methodologic issues. Dr. Steven Piantadosi is professor and director of oncology biostatistics at the Johns Hopkins University School of Medicine. He's here to discuss these issues. I think significantly, Dr. Piantadosi was the statistician responsible for the analysis of the original 8802 Gliadel wafer study in the GBM recurrent patients, is an author on the Lancet publication of those results.

Dr. Piantadosi?

DR. PIANTADOSI: Thank you, Dr. Hilt.

I'm going to discuss several of the methodologic issues that have arisen in the analysis and review of the current trial, and an outline of those points is on this slide. First I'll run over quickly the key design features of the study that were incorporated to eliminate or reduce bias in the overall estimate of the treatment effect.

I'll discuss some of the concerns about pre-specification of analyses. One of the more

contentious points in dealing with the review of this product has been the particular use of stratification, both in the design of the trial and the way that the sponsor has analyzed the data, and I will review fairly extensively our approach to that.

Then a final issue has been reassurance that significant strong prognostic factors have been adequately controlled and are not influencing the estimated treatment effect, and I'll discuss our approach to that.

All of the analyses that I present to you in the next few minutes -- and it's only a preview of the thorough analyses of the trial -- are based on the intention to treat population, and everything that I discuss will have been pre-specified in the study protocol.

As a review of the bias-reducing features in the trial the study, as you've heard, was a placebo-controlled double-masked study; somewhat unusual in oncology and certainly in surgical therapies, but the norm for the best, least-influenced estimate of treatment effect that we know how to design.

The original design of the study called

for a stratified block randomization within study center, and this in fact was implemented. The study also, as a result of this block by study country, because the centers are nested within countries. So the blocking and stratification within center also induces a similar blocking and stratification within country. This will be important later.

The study was not blocked or stratified by either histologic type, age, or Karnofsky performance score, and this is at variance with the suggestion in the FDA review document on page 37, which was not correct. The only blocking and stratifying characteristics were center and country. All the analyses, as I've said, were pre-specified in the study protocol to reassure us about control of type I error.

In the statistical analysis plan, the primary outcome variable was overall survival estimated by the Kaplan-Meier method. Treatment differences were to be assessed for statistical significance using the log-rank test and control of prognostic factors using the proportional hazards model.

One point of contention about the log-rank

test is whether or not that test should be stratified during the analysis. As I've emphasized already, this study was blocked and stratified both by center and country, and based on that design one would expect to employ a stratified log-rank test in the analysis. Literally, the protocol did not use the word stratified, nor did it use the word unstratified. I'll come back to this in a moment.

The pre-specified covariates based on what you heard epidemiologically and also from other studies include age, Karnofsky performance score, and tumor type. These are perennially observed to be clinically significant and statistically significant covariates in these cohorts. But also country of treatment was identified prospectively as a possible prognostic factor that needed to be controlled.

It's important to note that all of these factors, including country of treatment, induce variation in the outcome that is larger than the treatment effect. I'll show you that in a moment. Therefore, it's absolutely critical that one be able to control these effects, and have control over these effects, to guarantee that the estimated risk ratio is not unduly influenced by them.

A brief sketch of my approach to the analysis. I'm the person who has sort of crafted the basic approach here and conveyed that to the company. Initially, all the analyses were conducted by me personally. I reviewed the statistical analysis plan and the protocol before acquiring data from the sponsor and formed and impression as to what the proper analysis of this trial should be. I had no contact with the sponsor prior to transmitting some of the results to them.

My initial analysis used a stratified by country, as I've indicated, log-rank test, and I combined countries that were low accruers. Some had only one or two or three patients accrued. I combined them into a common group called an other country, if you will, and used that as one of the strata.

There were no post-hoc analyses conducted by me and no post-hoc analyses are presented today to base product approval on.

Now the stratified analysis, as I say, is motivated by the following reasoning. First, there's an explicit acknowledgement in the way that the study was written, the way that the randomization was performed, that center was an

extraneous source of variation that needed to be controlled. The use of block stratified randomization in a multicenter trial is absolutely off-the-shelf, standard approach. Based on that, one would expect the analysis to be stratified, and the analysis statistic to be stratified in the same way that the randomization was performed.

Now it is possible mechanically to do blocking and stratifying in the design and randomization and not stratify the test statistic, and vice versa. It's possible to do simple randomization and use a stratified test statistic. Mechanically there's no problem. But the optimal control over this extraneous source of variation comes from doing both. There is considerable discussion in the methodologic literature to support this, and this slide contains some of those discussions.

The first point is that treating known sources of variability as unknown sources of noise is really to be avoided, and there are several good papers on this. The one by Fleiss from 1986 in Controlled Clinical Trials is a pretty good one. Rich Simon, who was the former statistician on this very committee has written extensively about this

topic and certainly supports this perspective.

The second point about stratification is that over-stratification is also to be avoided. In the extreme, over-stratification is equivalent to no stratification at all. Imagine, if you will for a moment, a case where the study is stratified to that point that each patient enters his or her own stratum. This would be completely equivalent to simple randomization.

So too much stratification is also to be avoided, and limited stratification is in fact the strategy to be sought because it increases the sensitivity of the study by controlling the extraneous source of variation. So I'd refer any interested parties to these and other publications to support the general approach that I took to the analysis of this trial. There is, as far as I'm aware, no support in the clinical trials or statistical literature for blocking and stratifying the trial and then ignoring that randomization constraint during the analysis of the study.

So as I've indicated, randomization induces blocking and stratifying within country. It's probably the case that treatment practices vary more from country to country than they do from

center to center within country, and country has already been identified in the statistical analysis plan as an extraneous source of variation that needed to be controlled. Then finally I would point out that the nearly 40 study centers probably amounts to over-stratification if one were to use that as the stratifying level.

So it's my belief that stratification at the country level appropriately controls the source of variation and has a high degree of fidelity to the way that this study was designed and conducted.

This slide shows the results of the placebo survival in each of the countries in the study. You can see that there's a fair amount of heterogeneity here, and that heterogeneity is in excess of what one sees for the actual treatment effect, indicating again the importance of controlling country as a source of variation to be sure that you have an accurate estimate of the overall treatment effect.

The next slide shows what I think is a more informative and useful view of the study results. The top two-thirds of the slide show the results, the estimated hazard ratio within each study center. Here you can see a listing of all

the study centers, and the chart done in a meta-analysis style makes very clear the two sources of variability that one needs to cope with in analyzing the trial.

The first source of variability is typified by the approximate 95 percent confidence intervals around the hazard ratio estimates within each center. When the centers are small, as indicated by the smaller dots -- the dots or diamonds indicating the hazard ratios are drawn roughly in proportion to the number of patients accrued at that center. For the smaller centers you can see very broad confidence intervals. For the larger accruing centers, somewhat narrow intervals.

But the second source of variation, apart from person to person, is the variation from center to center, as the dots appear to be varying around this line of equivalence or no treatment effect.

In the bottom third of the picture you can see what happens to these estimates when a level of stratification taken at the country is used. Now the dots are larger because the country aggregates are larger than the clinic aggregates. The 95 percent confidence intervals are somewhat narrower

than they are for individual centers.

There's a general consistency of effect with the estimated hazard ratios falling on the side of the line, the left side, which indicates benefit for Gliadel. In fact there's only a single country, Australia, where the estimated treatment effect lies slightly to the right of that line.

The large dot which is third from the bottom is the overall result for T-301 with its 95 percent confidence intervals nearly obscured by the size of the dot. But the overall estimated hazard ratio lies to the left of the line indicating a benefit for the study drug.

Now the issue of stratification does not affect the estimated hazard ratio. That's the same whether one uses a stratified test or not. The stratification merely changes the denominator of that test, or the variance, and that has a small, but from a regulatory perspective, important effect on the p-value. So this location of this dot would not change as a result of the use of stratification.

The second dot from the bottom, a large one, is the result from 8802, the randomized trial in recurrent patients. You can see immediately

from this view that the treatment effect, the magnitude of the treatment effect and the approximate significance level is the same in study 8802 as it is in the current trial.

Then the smaller dot at the bottom with 95 percent confidence intervals is the result from the 0190 trial. This trial, as you heard, was deemed too small. But in fact, the estimated benefit for Gliadel in this very consistent with the other evidence. You can see that the magnitude of that effect is approximately the same as it is from this country, which is Germany, and this country here which is the United States.

So it's very clear when you look at the overall randomized evidence for Gliadel that all of the trials, and in fact all of the sizeable aggregates are telling you the same thing about the estimated treatment effect.

DR. PIANTADOSI: The last point I want to deal with is prognostic factors and being certain that these are not responsible for spuriously creating a treatment effect. All of the factors that I'm going to discuss were identified a priori in the study protocol. We used a very systematic approach to assessing the importance of those

factors. The first step was to perform univariable regressions where we identified the strength and approximate statistical significance of those factors and took those that appeared to be strong and significant using a p-value cut off of .05 and put those factors into multivariable regressions to assess their joint effect and their joint independence. The technic used for that was a standard one of proportional hazards model. The treatment of prognostic factors in the briefing book on page 39 does not follow this kind of algorithmic approach and is somewhat misleading in my opinion. It doesn't represent an a priori specification of how these analysis factors should be treated or even what they were.

The next slide shows the results of the first step of this systematic approach which is the univariable regressions. Here you see the now familiar prognostic factors, Karnofsky performance score H and number of wafers implanted, which depends on the size of the tumor cavity and therefore is a crude surrogate for the size of the original tumor and histologic type. An important point here is to note that all of these factors are strong. The risk ratios from about 1.5 to twofold

are generally in excess of the magnitude of the overall hazard ratio for the therapy, again indicating the importance of controlling these factors and all of them are strongly significant at conventional levels.

The next slide shows the result of putting those identified factors into multivariable regression with treatment effect to test whether or not that effect is influenced by adjustment for these factors. Although these factors are not statistically significantly imbalanced in the treatment groups, because they are so strong they don't have to be imbalanced to a high degree to be able to influence the result. So it's absolutely important to conduct this kind of analysis even though the factors appear to be balanced. In any case you can see that the overall Gliadel effect representing a risk reduction of about 28 percent is preserved in the presence -- adjusted for, if you will, these risk factors. The histologic type is still a strong factor but not statistically significant in a multivariable regression at conventional levels but Karnofsky performance score and age remain both strong and statistically significant. The point is that this analysis is

convincing that these prognostic factors are not driving the treatment effect. Both the univariable and the multivariable regressions that I've shown you are stratified by country based on the argument that I made previously.

So in summary, the study provides by design an unbiased and fairly precise estimate of the overall treatment effect from a design and methodologic point of view it is adequate and controlled. All of the analyses that I've presented to you and that I've performed and discussed here are rigorously pre-specified in the study protocol. The use of stratification as proposed at the country level is correct and consistent with standard statistical practice. The treatment effect is clinically significant representing the risk reduction of about 30 percent and convincingly independent of the influence of strong prognostic factors. Thank you very much, Dr. Hilt.

DR. HILT: Thank you, Dr. Piantadosi. I'd like to proceed now with the analysis of the T-301 study. First, the efficacy slide shown here is for the primary pre-specified endpoint which is the Kaplan-Meier overall survival analysis in the true

patient population. Shown here Gliadel produces a risk reduction of 29 percent. Survival benefit is statistically significant with a p-value of .03 as Dr. Piantadosi has already outlined using logrank statistics stratified by country.

Another baseline prognostic factor is a difference significantly between the two groups. However, to control for the effects of chance and balances in these various prognostic factors analyses while performed using the Cox Proportional Hazards model when accounting for prognostic factors that have a clear impact on survival such as in this case age, Karnofsky score and tumor histology. The treatment factor range is significant with the tumor histology following out of the final model. The risk reduction from the Cox model is 28 percent risk reduction. Therefore, one does not diminish the treatment effect of Gliadel wafer after accounting for important prognostic factors.

Our conclusion then is that for the primary pre-specified endpoint in this trial the T-301 trial is positive. There is a substantial increase in survival produced by the use of Gliadel wafer at the time of initial surgery in patients of

primary malignant glioma. The treatment fact is significant without accounting for known prognostic factors or the risk reduction 29 percent if one accounts for prognostic factors with treatment effect remaining substantial the risk reduction of 28 percent is significant.

Now the statistical analysis plan specified that a sensitivity analysis being conducted to account for additional therapies administered to patients at the time of tumor relapse. It was noted that a much higher percentage of patients underwent re-operation for disease progression than originally projected based on the 0190 study where only one patient actually underwent re-operation for tumor relapse. In the T-301 study 66 of the 240 patients had re-operation for disease recurrence or progression. There were similar numbers of patients in both treatment groups undergoing this procedure and why the patients undergo this re-operation.

Physicians re-operate due to disease recurrence, to relieve symptoms or to prolong survival. Sensitivity analysis was performed to account for the effect of this re-operation on the results of the survival endpoint by censoring

patients alive at the time of this re-operation for disease progression. Such analysis would provide a more precise measurement of the glioma for treatment effect. In addition such an analysis will allow a direct comparison of 0190 and T-301 studies.

Shown here, if one looks at the Kaplan-Meier survival analysis in the attempt to treat population censoring patients alive at the time of re-operation for disease progression, one sees approximately a 3.4 month median survival benefit and a statistically survival benefit shown in this Kaplan-Meier analysis. This represents a risk reduction of 36 percent shown here. This analysis most closely approximates the condition of 0190 study where only one patient had re-operation progression and arguably most accurately demonstrates a treatment effect that is conferred by the Gliadel wafer treatment alone without the potential confounding effect of re-operation for disease progression.

I will now the review the pre-specified secondary endpoints in the trial. These include overall survival in the GBM population of patients and other clinically important endpoints. As I

previously indicated these include time to Karnofsky performance decline, time to nerve performance decline, progression-free survival and a quality of life evaluation. Now in the GBM sub-population of patients in this trial the survival in the Gliadel wafer treatment group was increased versus a placebo wafer treatment group.

The effect is very similar in magnitude to the effect observed in the overall population and represents a risk reduction of 24 percent and the p-value for this effect is .1 shown here. However, when this analysis accounts for the effects of age and Karnofsky performance status the treatment effect becomes significant with a p-value of 0.04 with age following out of the final model. The risk reduction is 24 percent and after accounting for the effects of important prognostic factors the risk reduction is 31 percent.

Now clinically important pre-specified endpoint in the trial was time to Karnofsky performed score decline. Kaplan-Meier analysis of this analysis is shown here. It demonstrates a statistically significant Gliadel wafer treatment effect in maintaining overall function as measured by the Karnofsky performance score. The risk

reduction with Gliadel wafer treatment is 26 percent as shown here. Thus, Gliadel wafer treated patients maintained a higher level of function for a longer period of time so patients not only survived longer but at a higher level of overall function.

The next slide shows the analysis of 11 different pre-specified neuro performance measures similar to the time to Karnofsky performance score decline the time to neuro performance measure was measured for each of these 11 different neuro performance measures. These measures did not differ between the two treatment groups at baseline. Now these measurements assess how long patients can maintain neurologic function before undergoing a decline.

The Gliadel treatment confers a statistically significant benefit in 10 of the 11 neuro performance measures as shown in the slide. In the one measure visual status where the effect was not statistically significant there was a trend favoring the Gliadel wafer treatment group. All of these analyses have been stratified by country. I'd like to show just a few of the Kaplan-Meier curves to illustrate some of the benefits in

important areas of neurologic function such as speech as shown on this curve. Treatment effect is highly significant in patients treated with Gliadel have a clear advantage over the placebo wafer treatment group. An important point to make here is that the difference is shown here in the groups of 13 weeks, which is over three months, between the Gliadel wafer and placebo wafer groups is substantial and it represents time during which the patient is functioning at their initial higher level for this particular analysis.

This overhead now shows the cranial nerve function and it demonstrates a similar type and magnitude of effect treatment benefit. This overhead shows motor function which, of course, is very important to overall patient functioning. There's approximately a 14-week difference in the median time to decline in the Gliadel patients versus the placebo patients. This effect is statistically and I would argue clinically significant. And finally, cerebella function demonstrates a similar treatment effect.

Analysis of all these neuro performance measures have demonstrated a statistically significant and clinically meaningful benefit of

Gliadel wafer versus the placebo except in the visual status measurements which favor the Gliadel wafer treatment group. These changes are not only significant, they're clinically meaningful to patients and physicians. Contrary to the FDA briefing document it is our position that adjustments for multiple comparisons were not required because these analyses were pre-specified and they're intended to be supportive in nature yet not the primary endpoint.

Finally, presenting an overview of safety of the Gliadel wafer in the primary malignant glioma treatment setting, the next two slides summarize the safety profile of Gliadel wafer in this treatment setting. Intracranial hypertension, shown here, was more frequent in the Gliadel wafer treatment group, 9.2 percent versus 1.7 percent. However, brain edema was not reported with an increased frequency. Intracranial hypertension was typically observed late at the time of tumor recurrence and not in direct reaction to the wafer implantation procedure. Of the 11 patients treated with Gliadel wafer who had the intracranial hypertension diagnosed, 11 patients, 9.2 percent, 10 of the 11 patients had intracranial hypertension

reported more than 200 days after the implantation of the wafer at the time of disease recurrence. Therefore, this adverse event is not likely associated directly with Gliadel wafer use.

CSF leak, as shown here, was reported in more patients in the Gliadel wafer treatment group than the placebo wafer treatment group, 5 percent versus 0.8 percent. However, CSF infections per se were not more common with Gliadel wafer treatment group. Convulsions and other healing abnormalities were not more common in the Gliadel wafer treatment group versus the placebo group in this study. These results are different than that observed with the Gliadel wafer in a recurrent surgery setting, so-called 8802 study where these adverse events were more frequent. That's the re-operation disease recurrence study.

Now, we believe it's important to continue to advise clinicians to monitor Gliadel wafer treated patients for cerebral edema, signs of increased intracranial hypertension. Consequently, aggressive steroid use is clearly warranted in this patient population. CSF leak, though it is uncommon may be more frequent in the Gliadel treated patients than in the placebo treated

patients. Attention to surgical technique to assure a water tight dural closure is important. Taking all these considerations together, the safety profile of the Gliadel wafer appears to be more -- to be acceptable and is more benign in the primary surgery setting than in the recurrent disease setting.

There are no differences in systemic adverse events nor laboratory abnormalities between the two treatment groups. Now shown here is a more detailed listing of specific neurologic adverse events that occurred in more than 5 percent of the patients in either of the treatment groups. There are no significant differences between the two treatment groups in these neurologic adverse events with the exception of intracranial hypertension shown here and I've already discussed that. Therefore, the safety profile of the two groups appear to be very similar. Now, more specifically the frequency of convulsions, including grand mal convulsions was not different in the two treatment groups. An additional analysis was conducted that assessed the time to first seizure in the two treatment groups. That also showed no difference between the two groups. Five patients of the

placebo group experienced convulsions within the first five days post-operatively compared to three in the Gliadel wafer treatment group.

Now, during the T-301 study data on specific healing abnormalities was collected based on the clinical experience of Gliadel wafer in the recurrent surgery setting. The first such analysis specifically analyzed here is fluid CSF or subdural collections. No differences were observed between the Gliadel and placebo treatment groups as far as the frequency or median duration of a specific healing abnormality. The next CSF leak did occur at increased frequency in the Gliadel versus placebo treatment groups as I've already noted. The Gliadel wafer does contain BSNU local delivery of BSNU into the brain parenchyma may have local effects including edema and if a water tight dural closure is not attained, possibly promote a CSF leak. Six patients in the Gliadel wafer treatment group versus one patient in the placebo group experienced this adverse event.

Next is wound dehiscence, wound break down or poor healing. Again, this did not show any differences between the two treatment groups. Finally, a subdural effusion, subglial or effusion

demonstrated a similar effect, there are no differences between the two treatment groups. Therefore, the only healing abnormality was observed to occur in a higher frequency in the Gliadel wafer treatment group was the CSF leak, six versus one.

Now the frequency of intracranial infections in the two groups as shown here is deep infections of abscess and meningitis. Overall the infection rate was five percent approximately and there were no differences between the two treatment groups. Now the question has been raised in the FDA briefing document as to whether or not Gliadel wafer treatment is safe based on a comparison of the placebo wafer treatment group. Specifically, a suggestion has been made that even though the frequency of post-operative seizures, infections, hemorrhage or stroke complications are similar in the two treatment groups in the T-301 trial.

Placebo group may not be a representative control group because it involves the implantation of a placebo wafer. Shown here are some data on large series of neurosurgical patients a complication rate on this slide for post-operative surgical infections. Now some of these data are

included in the FDA briefing document, Table 20 and some of data on this slide are actually additional, very relevant publications and data that were not included in that table.

In a European study of 2,944 patients undergoing craniotomy for a variety of conditions Kornack, et al. reported an overall wound infection rate of four percent, the deep wound infection rate in that study was about two and a half percent of all patients. Brell, et al., shown here, reported 200 consecutive patients undergoing craniotomy for glioma metastatic disease.

They noted an infection rate of five and a half percent, and the deep infection rate in that study was 3.5 percent of meningitis and abscess in patients. Tenney, et al., shown here, 251 patients undergoing craniotomy for tumor resection. The deep wound infection rate of abscess and meningitis in that study was six percent. Therefore, the frequency of post-craniotomy infections in the T-301 study conducted predominantly within the European union is similar to a large EEU study of 2,944 patients conducted by Kornack, et al., and with these other published series. Secondly, seizures.

The number of studies that specifically the address the frequency of seizures after craniotomy for glioma, Cabantog, et al., shown here, this study was also in FDA briefing document Table 20, reported a post-operative seizure rate of only one percent. This was in 207 patients undergoing craniotomy for glioma. However, these patients were followed for only 30 days and in this study only patients whose seizure pattern changed preoperatively to post-operatively were included in this data. Therefore, patients with pre-operative seizures and post-operative seizures were not included in this tabulation.

Now this type of data is therefore not comparable to T-301 data where all seizures were recorded whether or not they differed from the pre-operative pattern. Now Brell, et al., as shown here, reported a post-operative "epilepsy" frequency of four percent in 200 patients. These are consecutive patients undergoing craniotomy for glioma or metastatic disease.

Now importantly, the definition of reported events in this series of patients are only those adverse events which qualify as "serious" adverse events and only those reported within 30

days of the surgery. Therefore, these adverse events by the typical definition had to be life threatening, cause hospitalization, birth defect, etcetera. This definition is obviously very different from the definition of an adverse event in the T-301 study where any seizure activity is reported. Pace, et al., in this slide shown here, report 119 patients undergoing craniotomy for glioma. The frequency of post-operative seizures varied in this study from 36 percent in the GBM patients to 83 percent in patients with lower grade tumors. Tanden, et al, reported 200 patients undergoing craniotomy for glioma. The frequency of post-operative seizures was 51 percent in this study.

Finally, a series of 65 consecutive patients by Moots, et al., shown here, they reported a recurrent post-operative seizure frequency of 32 percent. So, therefore, the frequency of post-operative seizures in the Gliadel and placebo patients in our study shown here appear to be similar to a number of published series of similar patients. So we, therefore, conclude that the frequency of infections and seizures after Gliadel wafers are similar in magnitude to the

frequency of those side effects after craniotomy in glioma patients.

The differences from the published series appear to be largely attributable to the different methods and definition used in collecting these adverse events, the only foundation Gliadel does not appear to confer at increased risk of side effects other than the CSF leak, which I've already discussed.

Finally, to summarize risk and benefits Gliadel wafer, first the safety. There was no evidence of early or more frequent seizures in Gliadel wafer treated patients in primary malignant glioma patient population as contrasted with our current GBM population. CSF leak, however, was more common in the Gliadel wafer treated patients versus the placebo wafer treated patients. There's no evidence of increase in intracranial infection or the healing abnormalities in the Gliadel wafer treated patients. Taking all these data together, safety profile of Gliadel wafer, the primary malignant glioma treatment setting appears to be acceptable.

To summarize the benefits of Gliadel wafer treatment, the use of Gliadel wafer in a larger

population of patients with newly diagnosed malignant glioma shows an increase in survival of patients treated with Gliadel compared to placebo wafer treatment. This effect is statistically significant, clinically meaningful as demonstrated by the results of two separate clinical studies now, the 190 study and the T-301 study. Currently this survival increase is accompanied by a maintenance of function in patients.

There is a delayed time to overall functional decline as measured by the Karnofsky performance score. The increase in survival is also accompanied by maintenance of good neurologic function. In 10 of the 11 pre-specified neuro performance measures Gliadel wafer treatment was superior to placebo wafer a treatment in delaying decline.

We think these results demonstrate the consistency of the Phase III Gliadel Wafer Studies. Two randomized, double-blind placebo controlled studies now demonstrate efficacy, acceptable safety in patients with malignant glioma undergoing primary surgery. The results of the 0190 and the T-301 trials in primary malignant glioma as well as the 8802 trial in recurrent malignant glioma

demonstrate the overall consistent efficacy of Gliadel wafer treatment in this patient population.

The risk reduction and confidence intervals for both the 190 and the T-301 studies are shown in this slide are the same data from the 8802 study. These data show then the three separate studies. Gliadel wafer has activity and produces a clinical of significant benefit in prolonging survival. The same analysis is shown here for the GBM subgroup of patients. The subgroup of patients as we know have the worst prognosis of all the brain tumor glioma patients. All these studies have now demonstrated the benefit. Therefore, Gliadel wafers having shown has significant efficacy in three randomized, placebo-controlled, double-blind studies in patients with malignant glioma, therefore, the benefit to risk ratio for Gliadel wafers in primary malignant glioma is favorable.

So finally, we therefore feel that the data support the following new indication for Gliadel wafer. Now this indication differs from the present indication as it provides Gliadel wafer at the time of initial surgery and indicates that Gliadel wafer maintains function inpatients.

Gliadel wafer is indicated for use as a treatment to significantly prolong survival and maintain overall function as measured by preservation of the Karnofsky performance status and neurologic function in patients with malignant glioma undergoing primary and over current surgical resection. I'd like to thank you very much and we'd be happy to attempt to answer any questions.

DR. NERENSTONE: Thank you. I'd like to open the floor to questions from the committee. I just want to remind you to try and keep it to questions to the sponsor and we'll save discussion for after the FDA presentation.

Questions from the Committee

DR. BUCKNER: I'd like to start out, please?

DR. NERENSTONE: Dr. Buckner?

DR. BUCKNER: A couple questions about study design first.

DR. HILT: Sure.

DR. BUCKNER: Why were patients less than -- over 65 excluded from the study?

DR. HILT: I'm not actually sure exactly why patients over 65 were excluded from the study designed clearly to have a worse prognosis than

younger patients. This study was conducted predominantly within the European unit was a feeling of the investigators that this type of patient would be the type that they felt they would wish to study and that they wished to enroll. Are there any other comments?

DR. BUCKNER: Do you believe that limiting to patients between 18 and 65 would have any impact on the labeling indications?

DR. HILT: Our feeling is that the drug appears to be very well tolerated and have an acceptable and relatively benign safety profile. There were a handful of patients in the trial who were actually over 65, four to six. Dr. Brem?

DR. BREM: The European investigators and Professor Westfall in Germany elected to exclude patients over 65 because they treat them differently there. They are less likely to operate on patients over age 65 and they felt that it's a much worse prognosis group and, therefore, they felt to compare apples to apples and limit it to the patients that they do full craniotomies on.

DR. BUCKNER: Another question. You've pointed out the importance of the prognostic variables because it can be a heterogenous group,

can you give us an idea of how you know your patients with glioblastoma actually had glioblastoma? In other words, what were your methods for reviewing the pathology?

DR. HILT: The way that this was done was that inter-operatively a frozen resection or a squash prep was done by the local pathologist who sent back up to the operating room the diagnosis, tentative diagnosis of malignant glioma or glioma. And the surgeon proceeded then, the patient was randomized through a placebo or Gliadel. The tissues were then sent to a central pathologist who reviewed them. And of course, the local pathologist did a formal view on fixed tissues. In the case of a disagreement between the local and the central pathologists where either one of them came up with a diagnosis of GBM but the other didn't, those cases were then forwarded to a referee pathologist. So if Dr. Dumondepor was the central pathologist and Dr. Reifenberger was the referee pathologist, so let's take the example where the local pathologist says it wasn't GBM, a central pathologist says it was GBM, the material was sent to the referee pathologist and he in a blinded manner made a separate read and it was the

best two out of three. So he was the final arbiter if you will. So that's how it was arrived at.

DR. BUCKNER: Do you have the data on the diagnosis of the central pathology reviewer, Dr. Dumondepor and could we see those data, please?

DR. HILT: I do not have the trial analyzed by that because we only included the final diagnosis is how the trial was analyzed.

DR. BUCKNER: How did you know that the central pathologist was not the correct pathologist?

DR. HILT: We have two diagnoses and we take this to a third clinician, in many cases the central pathologist was correct, perhaps in some cases the local pathologist was correct.

DR. BUCKNER: Do you have data available on the discordance rate between the central pathologist and the final pathologist?

DR. HILT: I have a slide that has the discordance rate with the local diagnosis versus the final diagnosis but I do not have the data you're talking about but I have that slide. If you want to look at that, I can provide that information? So in other words, that would be the local pathologist versus the final diagnosis after

this process ran its course.

DR. BUCKNER: I understand. Why don't you have discordance rates between your central pathologist who was considered to be your general expert and then your final pathologist who was also considered to be an expert to have their -- those would seem to me to be the most logical comparisons.

DR. HILT: There were roughly how many cases? I don't have that data handy. I don't have it here.

DR. BUCKNER: I respect the idea not to make comments at this point. However, those data were provided on the September 11th briefing package but not in this package and I wonder why? The reason it's important is there were substantial discrepancies.

DR. HILT: I believe there were about 30 cases that went out for this referee pathologist's review.

DR. BUCKNER: I think it would be very important for this committee to have access to those data because the percentage of patients with glioblastoma, between the randomized arms were substantially different than the final pathology

and also the percentage of patients with glioblastoma within each arm were different by central pathology compared with final pathology review.

DR. HILT: I think if you look at the percentage of patients in the two treatment arms after this entire process ran its course and we had three separate expert opinions, it's 106 versus 101 and I would argue that's not substantially different.

DR. NERENSTONE: Point of information. Will the FDA touch on that at all in your review?

DR. MARTIN: We did not bring a slide comparing local, central and final pathology. We'll be touching on the overall diagnosis as we referred to it.

DR. NERENSTONE: Dr. George?

MR. GEORGE: I have three questions for Dr. Piantadosi, if he's still here. These all relate to stratification. First point, first question is to make sure I understand it, the pre-specification issue or the new post-op analysis with respect to the stratification means that you determined before you saw the data what the stratification was going to be, that is that you

were going to use the stratified logrank test based on the stratification by country?

DR. PIANTADOSI: That's correct. I made that determination based on reading the design of the trial which was block and stratified by center and, therefore, by country.

MR. GEORGE: So in that sense it was pre-specified by you even though it wasn't in the protocol?

DR. PIANTADOSI: The protocol omitted the word stratified prior to the term logrank. I'm arguing that that's irrelevant.

MR. GEORGE: The other related issue is I looked back at the Lancet publication, the '82 data, in which case it clearly stated there was -- it was stratified by country and again omitted the word stratified logrank test in the analysis. Are you a recent convert to this approach?

DR. PIANTADOSI: No, I believe the Lancet publication, the proportional hazards model were in fact stratified by country. I may be misremembering but I'm pretty sure that was the case.

MR. GEORGE: In the methods it said stratification, the randomization was stratified by

country but I didn't detect it in the analysis but that may be. The third point I'd like to hear your comment on is the -- stratification in general when you're doing it this way is quite good because it increases efficiency but there's an issue with the logrank test and that is that any -- this is a statistical question, in each strata you're going to be losing information because of the -- any later observations beyond the last observed failure in a group is not going to contribute anything to the logrank statistic in particular if you have very few patients in the strata you may get -- you're definitely going to be losing some information. So the question at any given study is whether the loss of efficiency that way in any way has counteracted the other kind of gain of efficiency and, in fact, we don't know exactly which direction that would be. I'm just asking this as a general question, do you have any comments on that?

DR. PIANTADOSI: I think it's an argument -- I think you're absolutely correct, there is some small loss of information with the use of stratification in the way that you suggest. I think that it's an argument in favor of making the

strata fairly large, not too small, not too large, and not an argument against the use of stratification. If, as you suggest this is correct, and I believe it to be so, then stratification at the center level, despite the fact that that was literally the level of the randomization would involve more loss of information than lumping centers together in countries. If you play this game and look at different ways and levels of stratification it has some small but in my opinion non-definitive affect on the significance levels, suggesting that there is a little bit of information lost more or less depending on how you do this it, of course, doesn't change the overall treatment affect but does move the p-value slightly in either direction by a couple of percentage points.

DR. NERENSTONE: Dr. Blayney?

DR. BREM: Just to clarify about the Lancet study, which I'm the first author and the principle investigator, the --

DR. NERENSTONE: Could you please identify yourself?

DR. BREM: I'm sorry, Dr. Henry Brem. The Lancet study was primarily in the United States.

It was 27 medical centers. There were two centers in Canada, which is the only other country and they in aggregate brought in less than five patients divided between the two groups. So it was primarily -- there were two centers there less than five patients together. So of the 222, overwhelmingly it was a U.S. study.

DR. NERENSTONE: Dr. Blayney?

DR. BLAYNEY: The history of the sponsor, who first developed this drug and when was the trial conducted and when was the attempts -- and the data was presented to your statistical consult -- or the design was presented to your statistical consultant in what order of events?

DR. HILT: Sure. Dr. Smith, you want to comment on that? You're referring specifically to T-301 study now, correct?

DR. BLAYNEY: Yes.

DR. SMITH: I'll comment on the history of the product and then I'll let Dr. Hilt answer the second part of your question. Regarding the history of this product, it was originally developed by a company in Baltimore called Nova Pharmaceuticals. Nova Pharmaceuticals was bought by another company called Syos. Syos Nova, the

resulting the company formed a new company eventually called Guilford Pharmaceutical. And as a part of the capitalization of that company we acquired the rights to Gliadel wafer. On the evening of the 1996 ODAC panel where we received the initial approval, we signed a licensing agreement with what was then Roanpulankror Pharmaceuticals who designed and conducted the Phase III trial T-301. We paid half of the cost of the conduct of the Phase III trial but the responsibility for the conduct of the trial was initial Roanpulankror who was then bought by Hoechst Merian Racel forming Aventis Pharmaceuticals. Very recently, last year we reacquired the rights to Gliadel wafer from Aventis and have been responsible for the final analysis of the T-301 trial and the submission and all the regulatory responsibilities that come with making a submission to the Food and Drug Administration in April of this year for first-line therapy approval.

DR. BLAYNEY: And the timing of the presentation of your study design to your statistical consultants, was that before or after the study was completed?

DR. PIANTADOSI: I acquired the data only

after the study was completed and the legal agreements were signed for reacquiring of the product by Guilford. I believe I had seen the study protocol in the ASP technically prior to that but approximately the same time. I've had a long-standing, intermittent relationship with both Syos Nova and Guilford Pharmaceutical going back some number of years that I can't count at the present moment. So when they became aware that there might be need for analysis of a Phase III trial with Gliadel, I was contacted and I indicated my availability to the company to do those analyses.

DR. BLAYNEY: So that's a little different than I think trials are usually designed in that the study statistical analysis as specified as part of the study design.

DR. PIANTADOSI: This study analysis was in fact specified both in the protocol and the ASP, that's the document that I work from in performing my analysis. I was not the statistician of record, however, in the design or conduct of the trial, that's correct.

DR. HILT: The statistical analysis plan was drawn up, of course, well in advance -- was during the trial, in fact, during the beginning the

first half of the trial we commented to Aventis from Guilford that Aventis really had the primary need in finishing that document but the statistical analysis plan was finalized and had been in fact submitted to the FDA well in advance of the end of the trial. I want to make clear that point.

DR. BLAYNEY: I'd like to switch gears to another question. Two things. Why do you think the curves of survival separate at about eight or nine months?

DR. HILT: I think that many of the patients do quite well for a period of time and then the tumor relapses. So I think what this treatment might be doing is delaying the relapse in some fraction of those patients.

DR. BLAYNEY: In patients who were re-operated, were any of them re-implanted with your wafer?

DR. HILT: Yes, there were two patients who had Gliadel wafer at the time of re-operation for disease progression and there was in either group. So there were 36 patients in the Gliadel group who had mean time to re-implantation and re-operation of 260 days and then 30 patients in the placebo group who had mean time re-operation of

disease progression of 213 days and two of those received Gliadel, one in each group.

DR. BLAYNEY: What happens to the wafer at that point? Is it still there?

DR. HILT: We've done actually two patients, not in this study but previously, wafer remnants have been retrieved and analyzed analytically chemically and most of it is in fact water. There is a small number of polymer monomers mostly and there appears to be sort of a diaphanous structure mostly in water and polymer monomers. Dr. Brem is very informed on this subject if there are other -- he could perhaps elaborate on that if you want more detail.

DR. BLAYNEY: That's fine.

DR. NERENSTONE: Why don't we move on?

DR. BLAYNEY: The last question I have is looking at page 64 on your Karnofsky performance status decline and you picked preservation in your labeling indication, these curves look like they go straight --

DR. HILT: Now that is time to decline, that curve. So what that shows is the time, it's a Kaplan-Meier analysis of the time to when patients then have a decline in their Karnofsky score. So

we would argue that it is a maintenance of the initial -- or the effect of maintaining the initial Karnofsky score before it declines to a lower level.

DR. BLAYNEY: Thank you.

DR. NERENSTONE: Dr. Temple?

DR. TEMPLE: You've said, and it's obviously important, because the analyses vary depending on which one you use but the clear statistical analysis was presented to us well before the study was unblinded. Did that refer to what strata were going to be used, whether it was going to be country or site or any of those things?

DR. HILT: I was referring to the fact that the -- I think Dr. Piantadosi can comment on that, but I was referring to the fact that the statistical analysis plan was provided to the FDA well in advance of the completion of the trial. I was trying to make the point that the inference perhaps was there that the statistical analysis plan was written after the trial and I wanted to clarify that point. Now in the statistical --

DR. TEMPLE: Yes, but you didn't clarify it. You didn't you say what --

DR. HILT: I think the issue is that the,

as Dr. Piantadosi outlined, center was not specified in the analytical plan country-wise. If you look at the analytical plan, at the variables that were to be addressed in the Cox model that age, Karnofsky and country were identified, center was not.

DR. TEMPLE: Let's be sure about that. Steve said that he didn't look before at the data before he decided. That's not what I'm asking. I'm asking what you presented to us as your primary analytic plan? Did it say anything about stratification by country as the primary analysis?

DR. PIANTADOSI: Dr. Temple, the statistical analysis plan said stratification that randomization would be conducted by center. Did not say anything literally one way or another about stratification of the logrank statistic. What I'm saying is that I was the first person to look at the data when they were acquired by Guilford. I read the SAP and the protocol and I rendered some, admittedly some interpretation to the proper analysis given the design of the study primarily and the SAP. When you look now with the issue about whether we should be reporting a p-value of .03 or a p-value of .07 the omission or non-

omission of the word stratified becomes consequential and, therefore, there is ambiguity. All I'm telling you is that I was the first person to analyze the data. The first analysis that I personally did was to perform strata of moderate size based on country aggregates and analyze the data that way. The center was specified in the randomization. Country was specified as a variable of interest in the protocol. The sum total of that is an ambiguity about the way that the data were intended to be analyzed.

DR. TEMPLE: That's really the point I was trying to make. I don't doubt what you just said. I was also going to ask whether you sort of wrote that down anywhere so that, you know, you said, well, I looked at this and before I looked at the unblinded data I decided on this. But the company really didn't submit a plan that included all that as I understand it. It submitted an ambiguous plan.

DR. PIANTADOSI: Yes. The company asked me to adhere to the SAP and the protocol. Now I can't speak to when those were transmitted to the agency but what I had to go on was the SAP, the protocol and the data. And I believe based on my

analyzing these kind of data over and over again in many settings that I did what most people would do under the circumstance.

DR. TEMPLE: I'm not arguing for even which is right, but this is going to turn out to be important so it matters somewhat to know when everybody knew what was going to happen. I believe it's fair to say from what you've both said that the company submission to us was sort of silent on exactly how to do that, therefore, not particularly specified. That doesn't mean what was done was wrong, I'm not implying that in anyway.

DR. PIANTADOSI: I agree. What I understand was given to the agency, looked at now in retrospect, is in fact opened to interpretation about what could be called the protocol-specified analysis. I would argue, however, that in my position as a clinical trials methodologist I paid more attention to the design of the study for dictating the proper analysis rather than words on the page in the SAP. Imagine if you will that the SAP contained a technical error about what should be done. Surely nobody would expect us to adhere to that based on the design of its study.

DR. NERENSTONE: Dr. Lippman?

DR. TEMPLE: Those become major arguments when that happens.

DR. PIANTADOSI: I understand.

DR. HILT: On the screen is a -- both from the statistical analysis plan and also from an FDA review of a version of the statistical analysis plan dated August 22nd, 1997.

DR. PIANTADOSI: So my reading of this was that the randomization of this study was block and stratified both by center and country, that there was an a priori designation of country as an unwanted source of variation and looking at the accruals per center you recognize immediately that that level of stratification would be in the classic sense of the word over-stratification. So I did what I would consider to be obvious and reasonable in the way the data were analyzed. And I encouraged the company to agree with me, of course.

DR. NERENSTONE: Dr. Lippman?

DR. LIPPMAN: I have three questions. The first involves the placebo. I realize you must feel a little bit as though no good deed goes unpunished because from a trial point of view to have a placebo wafer in a treatment trial like this

so that you can maintain a double blind is I think very unusual for treatment trials. We do a lot of prevention. But so I think that it was a very sound design in that regard. But I also had a question about what the placebo wafer could be doing. I think you answered it nicely. I just wanted a clarification. For one of the slides you indicated that there was something like 10 to 20 percent severe -- between different studies, severe convulsions. My question really is, since I don't know this field, is that within on your slide 83 where you clearly show the overall rate of seizures is within what you'd expect in a control group that didn't get the wafer but when you break down those numbers here, slide 83, when you break down those by severe, does that also hold up as being -- in other words, the placebo wafer is consistent with the literature about it?

DR. HILT: What I'm showing here is a number of the studies above this first line I commented on the method by which they reported the seizures.

DR. LIPPMAN: My question is just of the seizures. These overall seizures presumably.

DR. HILT: Yes.

DR. LIPPMAN: You did a break down of the wafer by severe seizures and it was in the 10 to 20 percent range. Is that what these numbers that break down, the 36, 83 and 51 are intended?

DR. HILT: Yes.

DR. LIPPMAN: In the 20 percent range would be severe?

DR. HILT: That is my experience, but I would ask either Dr. Hamilton or Dr. Brem to comment as well briefly.

DR. HAMILTON: We don't tend to really group -- when you're talking severe if you mean a grand mal seizure as opposed to a petit mal seizure, we don't really tend to group them that way. What you're looking at here in order to have a comparable rate would be some of the glioma cases, these are glioma craniotomies which are a 10 with about a 20 to 50 percent convulsion rate. Now that includes all clinical convulsions. So that would be petit mal as well as grand mal and that does not differentiate between patients -- this includes patients who had seizures and developed seizures again post-operatively and does not differentiate if a patient has a more severe seizure disorder after surgery.

DR. LIPPMAN: I didn't make up that category. If you look at page 75, you have convulsions severe, so it means something, on page 75 of the T-301 study. And it ranges between the placebo is 20 percent but between 11.7 and 20 percent. All I'm asking is is that rate, how ever you clarify it or classify it severe consistent with --

DR. HAMILTON: Yes, that is comparable with the normal, traditional non-Gliadel wafer glioma craniotomies or tumor craniotomies.

DR. LIPPMAN: Thank you. The second question has to do with the intracranial hypertension. You indicated that there was a difference between the groups and you pointed out that's very late. I can't remember if you said it was 100 days or 200 days later.

DR. HILT: 200.

DR. LIPPMAN: So probably not related to the Gliadel but there was a difference between the Gliadel and the placebo.

DR. HILT: Eleven versus two.

DR. LIPPMAN: The question is do you think that's real or is it just a small numbers phenomenon because something must explain it if

it's -- in other words, do you have an explanation for that?

DR. BREM: I think it's not real from my clinical opinion. I think estimating even what's called cerebral hypertension, the clinical definition is so vague that it's a matter of whether the clinicians estimated that that was one of the issues with those patients. Any patient has a recurrence, virtually every patient has a recurrence, which virtually all of these patients eventually do in both groups, is going to have cerebral hypertension.

DR. LIPPMAN: I just wondered why it was different between the arms even though it was late but that's --

DR. BREM: It comes out as a difference but I can't think that is possible of being clinical significant.

DR. LIPPMAN: Small numbers probably.

DR. HILT: The other point I would make is that if you look at the frequency of cerebral edema, they're identical in the two groups so the consequences of having an increase in pressure, i.e. cerebral edema were not different.

DR. LIPPMAN: My third and last question

has to do with the 0190 study, and I realize that this is just supportive and so easy to look at differently so again I like the T-301 study. I thought it was convincing but I did have a question. It is a small study of 32 patients. I wondered whether there was a pre-specified sample size and what it was. Was it closed early because of the differences?

DR. HILT: Yes, the pre-specified sample size was 100 patients and Dr. Smith outlined the rather circuitous history of Gliadel through the various companies that have owned it and the reason that the trial was truncated at 32 patients is that material for the trial to continue was no longer available since Syos Nova were no longer making it. So unfortunately the trial had to be stopped in midstream very unfortunately.

DR. NERENSTONE: Dr. Albain?

DR. ALBAIN: This is for Dr. Piantadosi again. I'm making this point now just because of fear with airport schedules that the vote may occur when most of us have left, I hope not but just in case. I wanted your view of the slide 88, if you could put that back up please, and question for you. Whether we talk about using the adjusted

logrank for center or not, whether the p-value is .03 or .07, the effect that you're seeing here strikes me as highly consistent across whatever trial is done with this age but there's something going on. And I want you to put your statistician's hat on, could this still be play of chance or do you think there really is even if that p-value isn't quite .05, depending on how you view that adjustment?

DR. PIANTADOSI: I think it's very unlikely to be play of chance. You see a consistency in all the randomized evidence, which I tried to show on my quasi meta-analysis slide where all of the risk ratios are at this level of about 28 to 30 percent risk reduction or lower. We can quibble very much over what really is the correct type one error level for actually any of these trials for that matter. The 8802 study had some issues because of adjustment or not. This study has some issues because of stratification or not. The Scandinavian trial has some issues about whether the study was stopped appropriately or not. But my personal perspective is that this is a very real risk reduction. It's not a home run but it's clinically significant. If I were a patient with

brain tumors, I would find it to be an important effect. And quite honestly, I don't care personally whether you take the p-value to be .03 or .08. If you tell me that refuse to expand the indication of this drug into this population based on this kind of evidence and experience because you think the type one error rate is seven percent rather than five percent, that's not the kind of game that I would seek to play. I think that it represents a very real risk reduction and that's really where the emphasis ought to be.

DR. NERENSTONE: Dr. Fine?

DR. FINE: Along those same lines, I take what you say -- I actually agree also about playing the p-value bit. You do make a, when you talk about the decision to stratify center -- country versus center you make several categorical statements, and one of which I think is key is that you say that there is more likely or there's more variance by country in patient care than there is by center. I was just wondering whether you have any data to actually back that up?

DR. PIANTADOSI: No, that's pure instinct, Dr. Fine. I have no data to back that up. I think it depends on your view of the world. I think if

you're talking about the United States, you would look from center to center. We're a large country and we have varyingly trained oncologists. We would probably see quite a lot of heterogeneity from center to center. I think if you look at Europe, and I have very much more limited experience in trials in Europe than in the U.S., you also see similar center to center variation but in a relatively smaller country with fewer centers. I think there is more likely to be homogeneity across the centers than there is from country to country. That's a remark of instinct. I have no data or fact to back that up.

DR. FINE: Just another, as long as you're up there, statistical issue. Obviously, you've identified and controlled for the measure known prognostic parameters. There were two which I didn't see necessarily adjusted for. One was the issue of the extent of resection. As you know there have been a number of studies that have suggested from randomized trials that post-operative residual tumor is a prognostic factor, and in fact there was a slight difference in favor of Gliadel for gross total resection, 37 and a half percent versus 31 percent. So did you look at

that? And the other variable, if you could comment on is that the result rather large leeway, at least by U.S. standards and what type of radiation patients received and that they could receive between 5,500 and 6,000 centigrade RTOG has adjusted in all their databases that there is a dose/survival relationship in this disease. So the question is have you looked at the number of patients that got the lower end of the radiation scale versus the high end as a potential confounding variable?

DR. PIANTADOSI: Howard, I'll answer the second half of that first. The short answer for me personally is no. I did not have those data. I don't know whether those data on radiation dose exists or not. I'm aware that in some prognostic factor studies of brain tumors that dose of radiation is an important prognostic factor but I have not and cannot analyze data I don't have. With regard to extent of resection there's some information on that on this slide. You can see here a multivariable, I believe this is a multivariable analysis stratified by country. You can see percent resected here. This looks like a risk ratio that's very close to one but it's

probably close to per percent. So you have to think about this compounded over say 75 versus 50 percent and that kind of thing. P-value is marginally significant at conventional levels but really the important thing is that when you account for that as well as Karnofsky and age you see the same risk reduction that we're seeing all along. So in fact that variable, which is probably very strongly correlated or surrogated with some other predictor variables on the data set is not responsible for the putative treatment of that.

DR. FINE: The final question relates a little bit to Dr. Buckner's questions relative to ultimate, if it got to that point of labeling. It has to do with eligibility criteria, and maybe Henry can speak to that, of this trial, and that you gave us the rough eligibility, supercontorial gliomas with age cut-off but I do know that for many local therapy studies and so forth there are very significant exclusion criteria such as tumor invading the ventricle tumor involving corpus callosum. Were those exclusion criteria in this trial?

DR. HILT: Yes, the patient had to have unilateral tumor, could not have extension into the

corpus callosum or to the contralateral hemisphere.

DR. FINE: Or brain stem.

DR. HILT: Or brain stem, yes.

DR. FINE: So will those go into a labeling? Because that includes a large percentage of patients, if you know.

DR. HILT: I do not think that they are -- given the safety profile of the drug, I do not, myself see the logic of that. Maybe Dr. Brem will comment or Henry Friedman? Dr. Friedman?

DR. FRIEDMAN: Henry Friedman from Duke. I'd like to comment on three things actually that have been thrown as questions. Let's go in reverse order starting with Howard's question. I think that the appropriate use of this will be in patients who have essential major resections without extensive disease that is going elsewhere. I think in this country had we done the study I suspect that the restriction of age of 65 would not have been done. We have used since the paper was published Lancet we made a decision at Duke that Gliadel was standard of care for newly diagnosed patients and have put in well over 50 to 75 in the last number of years in newly diagnosed patients at any age. There is no difference in the toxicity

profile as a single institution admittedly limited experience in using this in patients over or under 65 as long as they had the kind of resection we're talking about.

I'd like to comment for a moment on Dr. Buckner's comments which are exactly on point. There is always the concern in glioblastoma trials that you're going to have a discordance of pathology. In fact, there have been some recent studies, some published, some unpublished which look at five senior pathologists, all leaders of their individual programs reviewing 100 cases of putative glioblastoma multiforme and only two-thirds could you get five out of five agreeing on the diagnosis. The rest are, the other third are split between four and one, two and three, three and two, etcetera so that the way we have approached, I think most groups have approached it, I'd be interested in how it's done at May or at the NIH, is that if you have a discordance between two pathologists, you seek a third opinion and you break the tie. Now, whether you want to use two of three, three of four, four of five, a majority, that's an arbitrary number but for us with the FDA or the NIH, NCI funded trials for patients with

glioblastoma multiforme, if we have a discordance, for example, treating an outside patient interpretation sent to Duke for a trial and we review it differently, it will get sent down to a third party and that is the tie breaker. It is not an easy diagnosis of glioblastoma as opposed to some of the other tumors the members of ODAC may be used to seeing.

Finally, with regard to the comments regarding was this really play of chance or is this really a true observation, speaking as a scientist first, I think the notion that in glioblastoma multiforme, we will see the kind of really I think explosive improvement in therapy such as with 571 and chronic myologic leukemia is remote. This is a very heterogenous disease. Perhaps no one in the world can speak to that better than Dr. Fine regarding the differences in the genetic composition patient to patient, tumor to tumor. So no one intervention is going to be the Holy Grail. There's no Glibac out there immediately obvious for GBM. So when you get any strategy that can increase survival in a realistic way and that strategy can be used with overlapping modalities such as the Temidar Gliadel we just published in

Neuro-oncology, the further trials that are going on with that, the combination of using Gliadel immediately with radiotherapy, you're beginning to make a -- chip away at the problem so to speak and ultimately going to result in an improvement. So at my institution where Gliadel is standard of care for newly diagnosed patient my biggest problem is not whether we want to use it, it's how we're going to get it paid for. There remains despite -- I think everyone at this table's concerns for the obvious are real problem with third-party payers who will say that if it's not labeled by the FDA for specification indication, it will not be paid for despite any published data. So for the patient advocate I would say if you want to see this technology out there and used in the newly diagnosed patient, which we dearly do, if it's not labeled, it may not be paid for, which obviously is going to prohibit its use.

DR. NERENSTONE: Thank you. Dr. Buckner?

DR. BUCKNER: I'd like to go on to some of the supportive data. I notice that one of your secondary endpoints was time to progression but you limit -- the presentation did not mention time to progression. Would you comment for the record?

DR. HILT: Sure. The record will now show that the time to progression of both treatment groups were equivalent. Could I have the criterion? Importantly, progression free survival in this study was not entirely an imaging or radiologic endpoint. It was a combination of a radiographic or imaging endpoint as is typical in these studies shown here in the bottom, and a clinical endpoint as well. Could you show me the break down please, the reason? And so what you see in the Gliadel and placebo groups here are the reasons for progression. If you sum up, there are 109 in each column, and if you sum up the patients who had progression due to an imaging criterion, it's roughly three-quarters of both groups. So the time to progression in this trial arguably is predominantly an imaging net criterion because three-quarters of the patients who did reach that endpoint had it based on an imaging study. So therefore, that's why this is obviously discordant with the Karnofsky time to progression and the neuro performance time to progression. Dr. Friedman, does that -- I mean that answers, I think answers your question?

DR. FRIEDMAN: Let's take it one step

further. This actually just I think supports a point that, Jan, you made at an ODAC meeting with tenzolomide where you gave a very articulate discussion of the problems associated with radiographic imaging as a parameter for progression free survival in patients with brain tumors. The point I think you made then was verified here in that I think what we're seeing may well be the consequences and changes on a scan of Gliadel when the relevant parameter is how they did clinically. So I think when you went on the Federal record back then you were right.

DR. BUCKNER: There can be multiple interpretations of why the scans look different. As a rule when scans look different from Gliadel then they subsequently improve over time if there is a beneficial effect. Have you followed up on that, do you have scans after progression that you have --

DR. HILT: No, I can't comment at all on that. Dr. Brem has had extensive clinical experience that he could comment on his clinical experience but in this trial that was not looked at at all.

DR. NERENSTONE: Maybe we'll keep it -- do

you want the answer or can we go on?

DR. HILT: That's fine.

DR. NERENSTONE: Because I think we should really stay focused on the clinical trial that we're being asked to evaluate.

DR. BUCKNER: I just have I think two relatively quick questions. How was neurologic status assessed? Was it the impression of the clinician --

DR. HILT: It was --

DR. BUCKNER: -- each of the 11 parameters?

DR. HILT: It was a neurologic examination where the same clinician looked at their previous exam and determined whether there was an objective change in their examination by normal, slightly abnormal, moderately abnormal, etcetera, shown here. So these I think have face validity because they're clinically observable to the clinician that changes in the neurologic exam.

DR. BUCKNER: Just one final question for Dr. Piantadosi on the supporting study, the small Finnish and Norwegian study. How would you describe the validity of a multivariate model of 32 patients with four variables and less than 32

events?

DR. PIANTADOSI: I'm similarly -- I take the gist of your question. I'm not totally comfortable with adjusting on two factors in 34 variables. I don't think one has to do that though to take the message of that trial. It is a small but unbiased estimate of the relative treatment effect of Gliadel. You saw from my meta-analysis slide that it's perfectly consistent with the magnitude and variation of results from other studies, other centers, those in Europe so I agree that that's not the preferred analysis but overall the risk ratio was significant in favor of Gliadel and strongly significant overall in the adjusted analysis.

DR. NERENSTONE: Dr. Moye?

DR. MOYE: Steve, I need to make sure I understand one of the comments you made in your slide. You said that all the analyses were rigorously prospectively specified. Does that mean that all of your analyses were rigorously prospectively specified by you before you carried them out?

DR. PIANTADOSI: No, that's my interpretation of what the SAP and the protocol

called for in terms of analyses. I've not gone on any fishing expeditions, for example. I have not gained any of the p-values with different strategies for multivariate adjustment with different strategies for stratification, with different outcomes or anything else. My read of the SAP and the protocol dictated those analyses that I did.

DR. MOYE: The second question I had was really just a response to something you said in response to Dr. Albain I think about this notion of .03 or .07. I think we all agree that's kind of trivial. I don't think that's the issue here though. The issue here is whether you can believe the estimates that these analyses provide. If the analyses are provided from well prospectively specified plans, then the estimates we have for relative risk, confidence intervals, p-values, are all accurate and precise and we can debate what they mean. However, if the analyses are developed from non-prospectively specified analysis plans or however we choose to define that today, then our estimates of p-values and confidence intervals are no longer trustworthy. So it's not the issue of .03 or .07. I'm sure we could all handle that

question very quickly. The question is whether we have estimates that are trustworthy or not.

DR. PIANTADOSI: I couldn't agree more. I've tried to convince you that the estimates, both the point estimates and relative magnitude of the treatment effect is trustworthy first. There's really nobody has pointed either in the questions or the substance of the trial to things that would bias the estimate of the treatment effect. To the contrary, we have about as objective a methodology and outcome as we can choose. This is a high standard foreign oncology and surgical trial, a definitive outcome masking the randomization and so on. So I think that the 30 percent risk reduction that you're seeing is our best unbiased estimate of the treatment effect. Furthermore, it's consistent with the other randomized evidence. What is the correct 95 percent confidence interval and consequently the correct p-value? There are circumstances where the way in which analyses are conducted will affect that. Obviously, that's a concern. I know that. I've sat where you're sitting now and I've debated these same issues in my mind. What I'm trying to convince you of is in fact that the estimates that I've provided you with

are the best that we can provide that they adhere to a pre-specified plan and consequently should be taken at face value. Dr. Albain's question went beyond that to say, well, if there is some reason why we should debate what the exact value, attach some consequence to that and that's what I was trying to answer there, that I would attach a very small consequence to which of the various debated type in error levels you choose to believe.

DR. NERENSTONE: Time is getting a little bit short. I have Mr. Ohye, Dr. Lippman, Dr. Rubinstein, Dr. Brawley and Dr. Lustig. Dr. Martin, do you need to respond?

DR. MARTIN: Dr. Piantadosi, I'm also going to be showing some slides of the results in the intent-to-treat population in the two trials that were submitted in 1996 and I'm sure we're going to confuse the committee because we have some different p-values. So I thought it was important to bring it up now even though we're running late. Specifically for on your page 88, Study 8802 has a p-value of .06. When we discussed this trial in 1996 it was the understanding that there were two primary endpoints, six month survival and overall survival and two analyses of those two endpoints,

both logrank and Wilcoxin that were incorporated into the protocol without ranking. Can you explain to me then this p-value of .06 which one that is of the four?

DR. PIANTADOSI: Yes. If you look at the protocol for Study 8802 and the protocol pre-specified analysis what you'll see is the Kaplan-Meier curve that Dr. Hilt showed in his presentation truncated at six months. That protocol pre-specified analysis was strongly significant using either the logrank test or the Wilcoxin test. One would not expect them to disagree since it's looking relatively early in the Kaplan-Meier curves. The p-value of .06 was one that I personally generated in analyzing the data after Dr. Brem requested that I be involved with the study. That .06 came from an overall analysis not only of all the data available at the time the study closed but some additional follow-ups that Syos Nova was able to obtain. And that .06 came from an overall logrank test looking at all the available follow-up.

The problem with the .06 and the issue that was addressed in the Lancet manuscript was that these strong prognostic factors, histologic

type, Karnofsky and age although apparently balanced in the treatment groups were slightly conspiring against Gliadel and an adjusted analysis showed, and what was presented in the Lancet paper, were predicted survival curves after adjustment showed that the risk ratio was probably more appropriately about a 30 percent risk reduction and the p-value as I recall was somewhere in the .02 to .03 range.

Again, for me it's the same issue whether you choose to accept the unadjusted, raw risk ratio estimate and p-value or the one based on multivariate adjustment. I don't care, the message is the same. There's about a 30 percent risk reduction in study 8802 for recurrent disease.

DR. NERENSTONE: Thank you. Mr. Ohye?

MR. OHYE: Actually my question has been answered. Thank you.

DR. NERENSTONE: Thank you. Dr. Lippman?

DR. LIPPMAN: Picking up just briefly on what Kathy mentioned because some of us won't be here later, just a quick comment. I think that it seems as though anyway these data are sliced or interpreted in terms of the statistical plan they show the same basic finding, and again reiterate

what's said about whether they hover around slightly above or below .05, and particularly this disease which lacks treatment and the toxicity data that we've seen, so I think it's very compelling and consistent but I do have one question. The issue of age versus performance status, we've debated that a bit at this meeting, and maybe Dr. Piantadosi can address this or whoever, but do you really feel that they're independent, that age is independent of PS?

DR. PIANTADOSI: No, literally. I don't think any of these prognostic factors are literally independent of one another but what the multivariable model allows you to do is to look for components of those factors that are in fact independent of one another. Typically what you see and what we saw here was that there's some modulation of the estimated risk ratios for each of those factors when they're considered jointly. And those relative hazards all tend to move toward the nul, toward 1.0. That in fact happened and the p-values tend to weaken slightly when they're considered jointly. The way that I interpreted the model is that what's left after the simultaneous fitting of those factors is the effect that is the

component of the effect that is independent of one another. So what you're left with in the adjusted model is the component of age that is independent of Karnofsky and so on.

DR. LIPPMAN: The reason I think it's important is when it comes to discussion of potential labeling and so on this issue that unfortunately the study didn't include older patients, that may be something that could be controlled for by performance status.

DR. NERENSTONE: Dr. Rubenstein?

MR. RUBENSTEIN: In your analysis in the book, I don't remember whether you covered it here, you gave the fully adjusted p-value stratified by country fully adjusted with age, performance status and tumor type. You gave those .03. On page 39 of the FDA book it's given as .1 and if you look carefully you see the difference is that the FDA has analyzed age as a continuous variable rather than a dichotomized variable. You analyze it as more than or equal to 60 versus less than 60. The question is when age was defined as a prognostic variable, was it defined dichotomously or was it defined as a prognostic variable to be used continuously?

DR. PIANTADOSI: It's my recollection, Larry, that the protocol didn't speak explicitly to that. I'm actually surprised at the premise of your question though. I didn't remember that particular analysis in the briefing document in that that was the only difference between those.

MR. RUBENSTEIN: It was on page 17 of your report.

DR. PIANTADOSI: I might ask Dr. Bordy or Dr. Hill from the company to refresh my memory on that. That's a sizable difference for the mere conversion of a continuous factor into a dichotomous one and I'm surprised is all I can say. There's not an issue of stratification by country in those two?

MR. RUBENSTEIN: No, they were both stratified by country I believe.

DR. NERENSTONE: Maybe we can go on. Dr. Brawley?

DR. BRAWLEY: A brief question. This actually revolves around one of Dr. Buckner's early points. What proportion or what number of patients underwent a resection and on either frozen section or squash section it was said that they glioblastoma and they were put into the trial or

randomized into the trial and then on further evaluation when permanent section was found it was some other tumor?

DR. HILT: Typically what is done during surgery is not the diagnosis of glioblastoma per se. It's either malignant glioma or not and that was the guidance that was given. If you can put back up the distribution of tumor types, that baseline, what you'll see is that the vast majority of patients had malignant gliomas of different grades. There were a handful of patients of nine I think that had other diagnoses such as astroblastoma, permanent neuroepidural tumor and there were a couple of patients who actually had metastatic disease so that the tissue sent down by the surgeon from the operating room, the provisional frozen diagnosis was "glioma" and only on the fixed tissue was the final diagnosis of the metastatic lesion diagnosed. So that this is the -- the surgeon does not have the luxury of a definitive diagnosis in the operating room. They have a provisional diagnosis. I think Dr. Brem will comment briefly on that.

DR. BREM: Very briefly. In terms of practical use of Gliadel, the standard approach

that we use and many other centers use, is that unless the pathologist says that it's a malignant primary brain tumor, Gliadel wouldn't be used. So the danger is not using it when in hindsight on the permanent sections it turns out to be a primary malignant tumor and it could have been used. I know my own experience which is several hundred patients with Gliadel, we've never made the error of placing it in a patient who doesn't have a primary malignant brain tumor. The distinction between the subtypes, whether it's an anaplastic oligo, whether it's a malignant glioma, anaplastic or GBM really sort of sorts itself out on the analyses after the permanent sections are in. Our pathologist, Peter Berger, who is reasonably good at this stuff won't even attempt to make those distinctions at the time we need it which is at the surgery. So that's sort of looking at a prognostic factor but not -- and all of the benefit from aggressive chemotherapeutic approaches.

DR. BRAWLEY: So the number of proportion of people who were treated inappropriately in this trial was -- well, I shouldn't say inappropriately but you know what I mean they --

DR. HILT: None of these patients -- all

of these patients had a malignant tumors. They're tumors of different types so when you look at the other -- we're not treating congenital malformations, etcetera. These are patients with different types of very esoteric, malignant gliomas and malignant tumors and the patients with metastases, the three patients, not four, three patients have brain metastases which looked at frozen section from surgical pathology like a glioma. So all of these patients have tumors.

DR. NERENSTONE: Mr. Lustig?

MR. LUSTIG: Just getting back to the issue of the post-surgical seizures. The comparative studies that you referenced, did any of those have in the study population the age limit that was in the Gliadel studies?

DR. HILT: I really do not know. I can't answer that for sure, to be honest with you. I can't recall. I have the papers over there. I could look afterwards but I can't tell you right now.

DR. NERENSTONE: Thank you. I'd like to thank everyone, and the sponsor. We're going to have a very brief break. I'd like everyone back at the table at 3:25.

[Recess.]

DR. NERENSTONE: Dr. Martin, if you'd like to get started.

FDA Presentation

DR. MARTIN: Thank you. Madam Chairman, members of the committee, ladies and gentleman, I would like to thank you for reconvening. As some of you know this application was scheduled to be presented on September 11th and we are grateful that you were willing to reconvene to give us a full hearing.

The presentation from the FDA is outlined on the slide and will consider some pertinent aspects of the regulatory history but many of these have been brought up already so I can skip. We will then hear a clinical and statistical commentary on the primary trial submitted to support this indication. Then I'll come back to summarize the review issues that face us and that lead us into the questions.

As you've heard, Gliadel is a marketed drug and currently it's indication is for not all recurrent malignant gliomas, a subset of patients with glioblastoma multiforme for whom surgical debulking and resection is indicated. The two

trials that supported this indication in 1996 have already been brought up so I will only remind you we'll be returning to them later as evidence for replicability for confirmatory evidence.

After the 1996 ODAC when the indication was not extended into the newly diagnosed population we had a meeting with the company. The major agreements are listed on this slide and included that a single trial could possibly support a new indication if it were multicenter with consistent results across the center and results were robust. The population of interest from both parties was the glioblastoma multiforme population, primarily because of the data from the North American trial that showed this appeared to be a more sensitive tumor.

It was, however, discussed that the histology would not be sufficiently well-defined prior to randomization to make this an intent-to-treat population. We weighed the pros and cons of a placebo wafer and that has already been discussed today so I won't belabor it. There was agreement to standardized subsequent treatments and to prospectively define local toxicities of interest.

While we agreed on those issues there were

some disagreements on how the sample size -- the size, what it was powered on the assumptions. The sample size was based on a 20 percent difference in a 12 month survival rate between the treatment arms. We commented on the protocol at that time that we expected that we'd be overly optimistic and did some modeling for the sponsor, that should the treatment difference only be 12.5 percent at that time the power drop to 53 percent.

The protocol proceeded without change. An amendment was submitted in 1999 enlarging the sample size from 200 to 240. This is when the independent data monitoring committee had reviewed the data in a blinded fashion and forwarded comments to the steering committee of the protocol that the hoped for surgery benefit of Gliadel of 20 percent, one year is probably unrealistic, a smaller but worthwhile benefit might be missed. At that point the sample size was increased and it was modeled that now an 18 percent difference between the arms would be detectable.

At this point I would like to introduce Dr. Shapiro who will start the clinical and statistical review.

DR. SHAPIRO: Thank you. Current

indications for the Gliadel wafer is different from the previous Gliadel approval in terms of the stage of the disease, newly diagnosed glioma versus recurrent, and patient population, intent to treat for this study and GBM subgroup for the previous application. Survival in the intent to treat population was the primary efficacy endpoint for this study. The secondary endpoints are listed on the slide.

In the statistical analysis plan GBM population was defined as speculation of interest for the treatment effect. The protocol did not rank the secondary endpoints and there was no adjustment in the statistical analysis plan for multiplicity.

The trial design, all patients were randomized to Gliadel or placebo group. Randomization was stratified by center. At the maximum surgical resection all patients were to receive limited field radiation therapy. Subsequently, all patients with the histological diagnosis of AOD were to receive six cycles of chemotherapy. No systemic chemotherapy was permitted for treatment of any tumor for patients with other histological diagnosis.

For study enrollment, a total of 240 patients were enrolled at 38 centers in 14 countries. The largest number of patients were accrued in seven centers in France and in five centers in Germany. In the United States only 12 patients were accrued in five centers.

The next two slides show the distribution of three known and accepted prognostic factors in newly diagnosed glioma. Patient's age and baseline KPS is reasonably balanced in both groups.

On tumor histology, we agree with the sponsor on the number of patients with GBM and the number of patients in non-GBM group. Our table differs from the sponsor's in three patients classified as other by the sponsor, shown on our slide by the actual histological diagnosis based on the assessment of the central pathologist. Overall there were slightly more patients with a favorable histology in the Gliadel than in the placebo, 17 and 13 patients, respectively.

Protocol-specified treatment included radiation therapy, chemotherapy, and other treatment for the disease progression. Standard radiation therapy was delivered to 78 percent of patients on the Gliadel group and to 80 percent of

patients on placebo. The remaining of the patients received either non-standard radiation therapy regimen or no radiation therapy.

Chemotherapy. This table summarizes all patients who received chemotherapy, both with a disease progression as well as for other histological diagnosis such as AOD and AOA. There were 11 patients with AOA and AOD. Although all patients did not receive chemotherapy as their protocol, the numbers of patients who did receive is balanced across the arms.

This slide presents additional treatment that could potentially impact survival. They are re-operation, with or without Gliadel reimplantation, and radiation. Overall there was difference only in two patients receiving additional treatment between both arms.

Efficacy results. Primary analysis for survival was to be conducted 12 months after the last patient has entered. A total of 88 patients in the Gliadel group and 93 patients in the placebo group died before the study cutoff group. In the Gliadel group, median survival was increased by two months compared to the placebo. The protocol and statistical analysis plan specified a log-rank test

for the primary analysis which did not reach statistical significance. The sponsor presented primary analysis by log-rank stratified by country which had a p-value of 0.03. This is one of our review issues and now will be discussed by Dr. Li in greater detail.

Thank you.

DR. LI: Thank you. I'm going to discuss the statistical issues in the primary efficacy analysis. The primary analysis proposed in the sponsor's protocol as well as in the statistical analysis plan was to compare the overall survival in the two treatment groups with their log-rank test. The log-rank test stratified each of the four prognostical covariates; i.e., Karnofsky performance score, age, tumor type, and the country, were to be performed as secondary analysis and is considered as supporting to the primary efficacy analysis. Supportive analysis is meant to strengthen the evidence provided by the primary analysis when the primary wins.

This is the resulting Kaplan-Meier survival codes for the study. The dotted curve is the Gliadel arm and the solid line is the placebo arm.

The protocol specified primary analysis result is summarized in this slide. There were 88 events in the Gliadel group and 93 events in the placebo-control group. The estimated median survival difference is about 2.3 months with a hazard ratio of 0.77 in favor of the Gliadel group. But there is no statistical significant difference between the two arms with the log-rank test p-value of 0.08. This p-value should be adjusted upwards since there was an interim sample size increase from 200 to 240.

The sponsor claimed a 23 percent or 29 percent risk reduction based upon the hazard ratio point estimate, but since the 95 percent confidence interval upper bound can not exceed one, statistically speaking, the evidence is insufficient to conclude that the risk in the treatment arm is lower than in the placebo arm at the 5 percent significant level.

As mentioned earlier, the supportive secondary analysis for the primary endpoint using a log-rank test stratified by the pre-specified prognostic covariates were performed. This slide shows the results. The log-rank test stratified by Karnofsky score has a p-value of 0.07, and a log-

rank test stratified by age has a p-value of 0.1, and by GBM type has a p-value of 0.14. The analysis stratified by center has a p-value of 0.07, but the center was not a pre-specified stratification variable. The only significant difference is a log-rank test stratified by country, which has a p-value of 0.03.

The analysis adjusting all pre-specified prognostic variables were performed as secondary analysis. Entries in this table are the p-values for the treatment effect. The sponsor commented on our analysis on page 39 of the briefing document and from which this slide came. We believe that step-wise selection procedure such as step-down procedure is not appropriate because, as Dr. Rich Simon commented in an ODAC meeting, p-value based upon the step-wise recreation is not interpretable.

When adjusting all of these covariates, no statistically significant treatment effect can be detected in three types of analysis. These analyses are supportive analysis. As I mentioned earlier, supportive analysis are used to strengthen the primary analysis results when the primary wins. There is a little bit of difference between the FDA's analysis and the sponsor's analysis, and for

this exposure analysis the age was treated as a continuous variable in the FDA's analysis while the sponsor cut the age to two categories, greater or equal to 60 and less than 60.

To summarize the survival analysis, all results of survival comparisons between the two arms are not statistically significant except to the analysis stratified by country. The sponsor presented the stratified by country analysis as the primary analysis and concluded significant survival benefit. The sponsor's argument is that stratified analysis is appropriate because randomization was stratified by country, and stratified by center analysis may cause overstratification.

Now we have two statistically related issues. The first issue is, should one use a stratified or non-stratified analysis? Which one is more appropriate? Our position is, either one is acceptable as long as you pre-specify one in the protocol. Retrospective selection is problematic because it will inflate type one error.

The secondary issue about stratification has been kind of resolved after discussing with the sponsor and we came to an agreement that the randomization was stratified by center, not

country. We can tell this by checking all 12 U.S. patients in all five U.S. sites. This is the randomization list. A fixed block size of four was used. If the country was a stratification factor, then the patients with similar dates should be classed in together.

For example, four patients entered the study in January, February, and June of '98 and patient ID with asterisks and italics, and patient ID 2005, 2013, 2021, and 2024 should be in the same block. But it's not the case here. We believe that randomization stratified by center may not necessarily result in a randomization sample in country.

If we believe a stratified analysis should be used, then according to the sponsor's argument the center should be the stratification factor. The result is similar to non-stratified log-rank test with a p-value of 0.07 as shown in this slide.

To conclude, protocol specified analysis for overall survival was not statistically significant with a p-value of 0.08. This p-value is subjected to an upward adjustment due to an interim sample size increase. The log-rank test stratified by center and all other stratified

analyses, which includes stratified by age, stratified by performance score, stratified by tumor type, or adjusting all pre-specified covariates are not statistically significant.

The sponsor's analysis, log-rank test stratified by country, one of the protocol pre-specified secondary analysis for survival is questionable as the primary analysis because, one, it is not pre-specified as the primary. Two, the result is not supported by secondary adjustment analysis. And three, if both stratified and non-stratified analysis had been pre-specified as part of the decision group then multiple analysis would be an issue and a certain upward adjustment is needed.

Dr. Shapiro will present the results for secondary endpoints.

DR. SHAPIRO: Thank you. In the statistical analysis plan, GBM subgroup was chosen as a population of main interest for treatment effect. Of the 240 patients enrolled, 207 carried the diagnosis of GBM. A total of 79 patients in the Gliadel group and 85 patients in the placebo group died before the study cutoff date.

Overall survival in this population

demonstrates a non-significant trend favoring the Gliadel group. The difference in the point estimate of median survival is two months. Statistical significance between the treatment arms was not shown by either stratified or non-stratified tests. A comparison of one-year survival rate in both the ITT and GBM subgroup appeared to favor Gliadel, but they're not statistically significant by log-rank non-stratified or stratified.

The sponsor's analysis of progression-free survival showed no difference between the treatment groups. FDA did not analyze the secondary endpoint. We consider progression-free survival difficult to assess in this patient population previously treated with surgery, radiation, or steroids.

Time to Karnofsky performance status deterioration was one of the three quality-of-life measures pre-specified in the protocol. In a non-stratified log-rank test this prognostic factor did not reach statistical significance. Time to KPS deterioration becomes statistically significant if the log-rank is stratified by country, not by center.

In assessing time to KPS deterioration, the sponsor counted death as an event. To assess the impact of death, the FDA performed an analysis by censoring patients who died. The log-rank test did not reach statistical significance in this case by any of the analyses.

Quality of life was also assessed by EORTC quality of life questionnaire 30 and brain cancer module, a validated 24-question instrument designed to be used in conjunction with quality of life. The primary QOL parameters pre-specified in the protocol was a measure of global health status based on questions number 29 and 30. There were no differences between the arms. However, it should be mentioned the study was not powered to show significant difference.

The sponsor presented a summary of data collected for the 11 pre-specified neuroperformance measures. These p-values are based on an analysis stratified by country. There appears to be consistency in outcome across these measures. However, we have concern on the assessment tool and statistical analysis.

With regard to the assessment tool, the categories for grading were tied to normal,

slightly abnormal, moderately abnormal, severely abnormal, not able to measure, and not done. Specific or objective criteria for choosing a category were not provided. Second, a change by one category counted as an event. And thirdly, death was counted as an event in this case as well. If death is censored rather than counted as an event, much of the data is lost.

For example, we censored death as an event. Only 11 patients showed a level of consciousness deterioration. Another example, the remaining 25 percent of patients in this -- after patients were censored for death, only 25 percent of patients in that category were able to be assessed for vital signs.

Finally, these results are not supported by findings in the parameters of KPS and QOl or adjusted for multiplicity. If we disregard issues of multiplicity and of the assessment of two for a moment and conduct an analysis where death is censored rather than counted as an event, the statistical significance is not apparent.

Safety results. In assessing safety results we will be focusing on death within the first 30 days of randomization as well as local

complications. The agency looks at death within 30 days of therapy as possibly related to therapy. The groups are balanced for systemic causes of death within 30 days; two in each arm. Only the Gliadel arm had death due to local complications such as cerebral hemorrhages. Local complications are presented on this slide.

We agree with the sponsor's assessment of occurrence of the most common local complications after wafer implantation. There are numerical differences in incidences of intracranial hypertension, CSF leak, and postoperative mortality in the Gliadel group. It is also possible that the assessment of risk may be underestimated since the control group is placebo wafer. We'll ask the committee to weigh the significance of these findings against any benefit when we conclude with the questions.

Now I'd like to turn the podium to Dr. Martin who will present the review issues. Thank you.

DR. MARTIN: Our usual requirement for evidence of drug efficacy is more than one adequate and well-controlled trial. However, the Modernization Act of 1997 specifically allows for

approval of a drug based on one clinical trial under certain circumstances, and especially if accompanied by other supportive evidence. However, reliance on a single trial generally is limited to situations in which a trial has demonstrated a clinically meaningful, statistically persuasive effect on an important endpoint such as survival.

The single trial paradigm, as demonstrated on this slide, starts with the necessity of an adequate and well-controlled trial. We will be asking the committee if this T-301 is considered adequate and well-controlled in light of the discussions that we've had about the pre-specification of the primary endpoint and multiplicity adjustments. We'll get into that a little bit later.

The single trial paradigm also mentions that the trial ought to be multicenter with consistency of results across those centers, consistency across subsets of patients, across secondary endpoints, and as I said, statistically persuasive.

It's not necessarily that we are a slave to the p-value of 0.05, but there's a general consensus that that level of significance in the